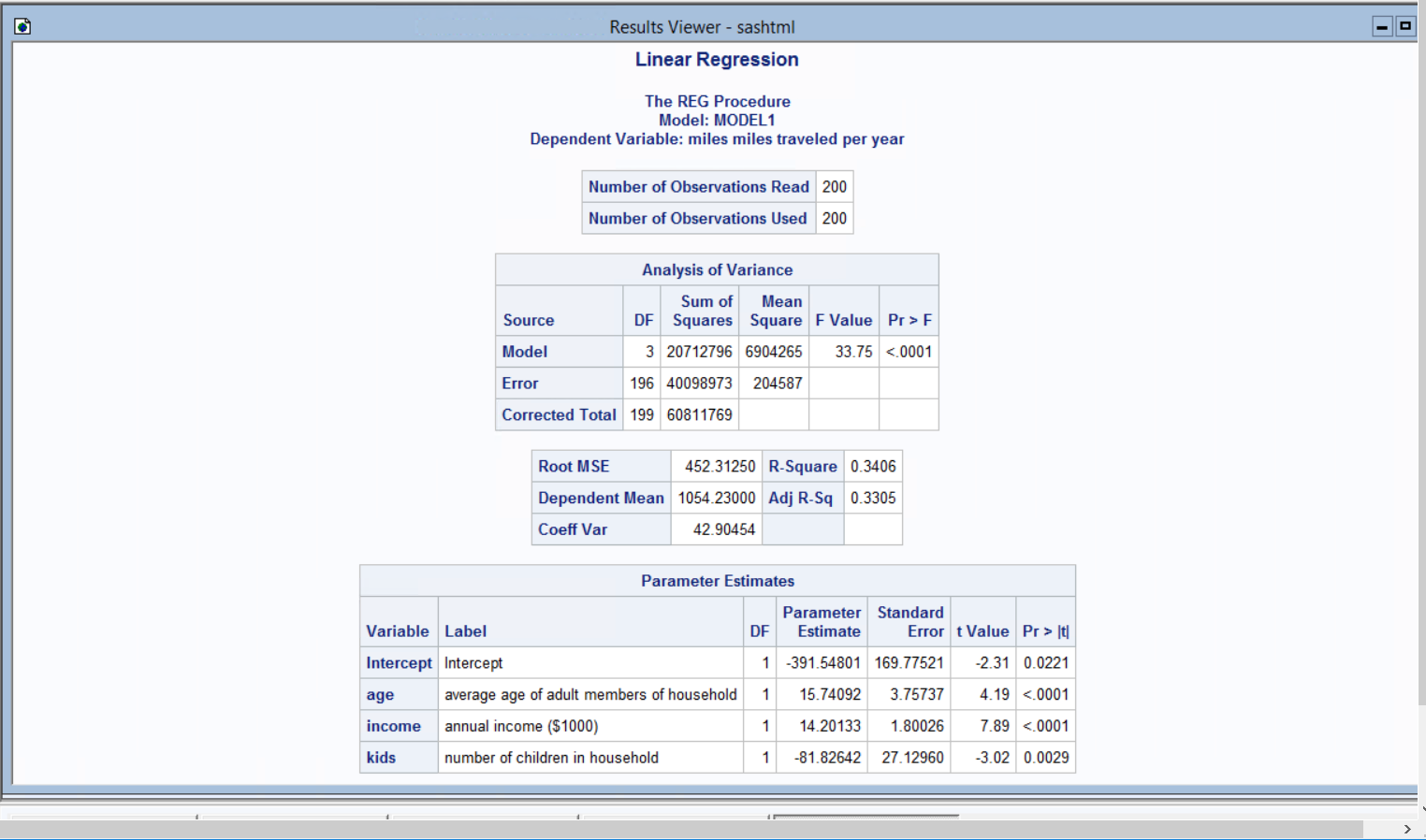
**Homework 3**

1. Use the data in vacation.dat and run a regression with Miles (miles travelled) as the dependent variable and income, age (average age of adult members), and kids as the independent variables.
2. **Run a regression model and interpret the coefficients. Comment on the model fit.**



For the vacation dataset we observed an R2 value of 34.06%. This means that 34.06% of variance in the dependent variable i.e. Miles is explained by three variables i.e. age, income and kids.

All the three variables (i.e. age, income and kids) have p-value < 0.05 and hence they all are significant to determine the miles.

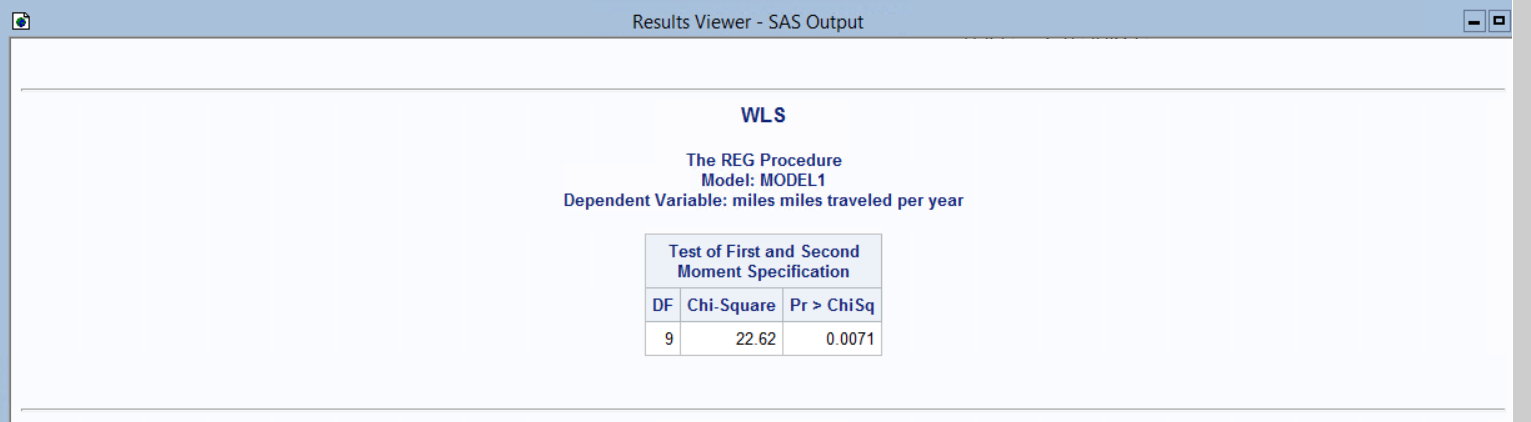
Age and Income coefficients have positive relation with dependent coefficient miles. So, if age increases by 1 year then there is an increase of 15.74 miles travelled. And if income increases by $1000 then there is an increase of 14.20 miles travelled. On other hand kids have a negative relation and every addition of a kid in the family decreases the miles driven by 81.82

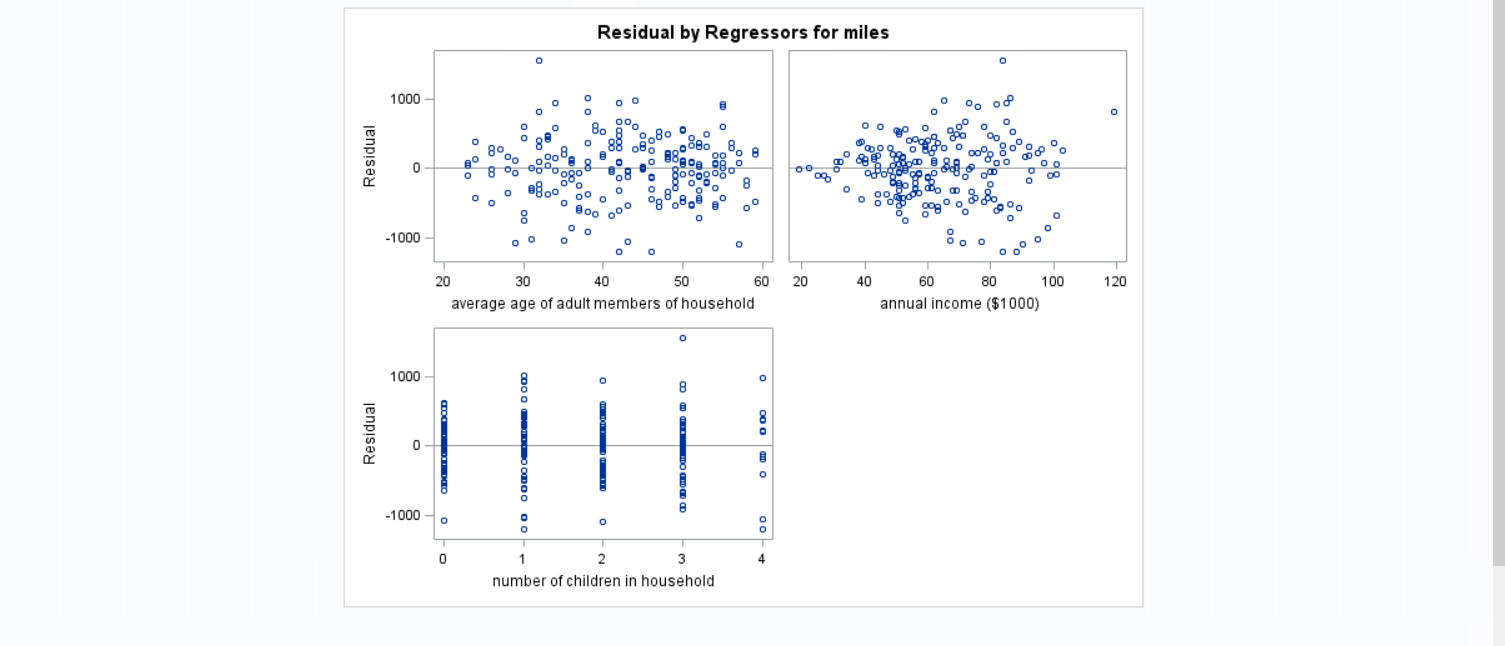
1. **Check whether there is heteroscedasticity in the model using White test.**

We considered following Null hypothesis:

H0 = There is no heteroscedasticity in the model

Ha= There is heteroskedasticity





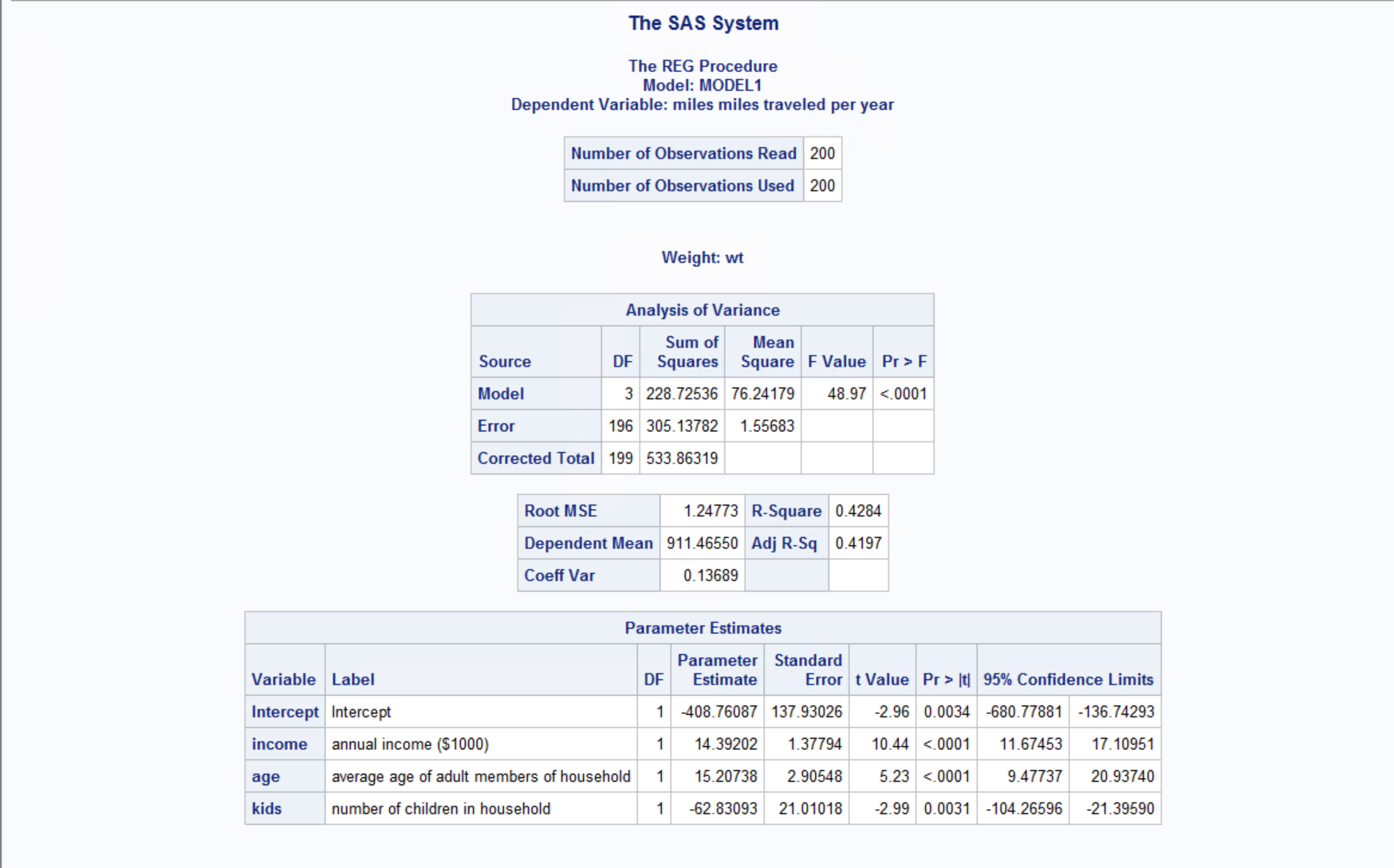
On running the white test using proc reg we got the output of p-value < 0.05 (p-value is 0.0071) from ChiSq test. So, we reject the Null Hypothesis. Hence, there is evidence of heteroscedasticity in the model.

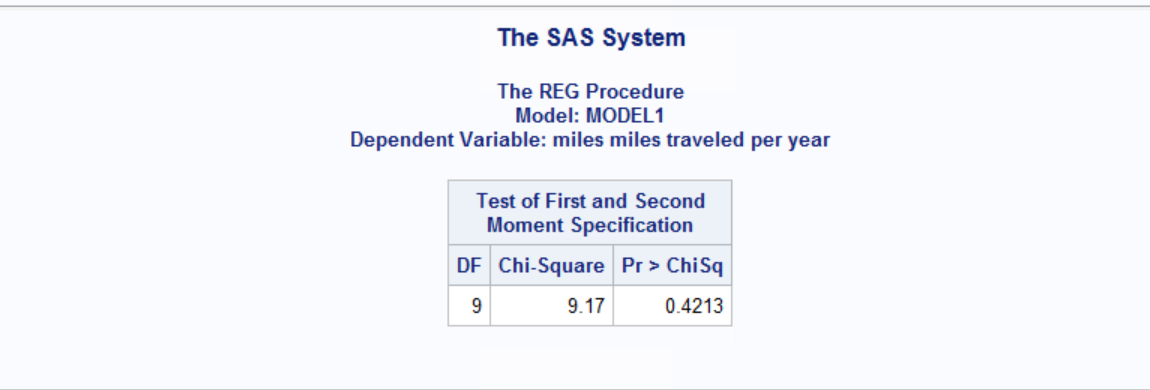
1. **Run a weighted Least squares (WLS) regression. Discuss your results in a paragraph. (Comment on model fit, significance of coefficients, and the effect of doing WLS.)**

WLS method is used when there is evidence of heteroscedasticity in the model. WLS is used to mitigate the problem caused by heteroscedasticity. WLS helps in correcting the error caused due to heteroscedasticity.

We defined the reciprocal of each variance as the weight and then use these weights in estimating a weighted least squares regression model.

The weighted least squares estimate is then estimated. Since each weight is inversely proportional to the error variance, it reflects the information in that observation. So, an observation with small error variance has a large weight since it contains relatively more information than an observation with large error variance (small weight).





We observe that the R2 value is 42.84% for a WLS model. The coefficients of WLS are nearly same as the ordinary least square estimates. All the three independent variables (i.e. age, income and kids) have p-value < 0.05 and hence they all are significant to determine the miles. The white test on the WLS also reveals that there no evidence of heteroskedasticity.

**Q2.**

**2.1 Using SAS and regression, estimate the Bass model. Save the regression parameters using option OUTEST. Find p, q, and M and compute peak sales and the time when that peak will occur.**

Regression model of sales is:

St = a + b Nt-1 + c (Nt-1)2

Where

a=pM

b=q-p

c=-q/M

By running the regression model, we get the following regression parameters:

a = 539.73

b = 0.310

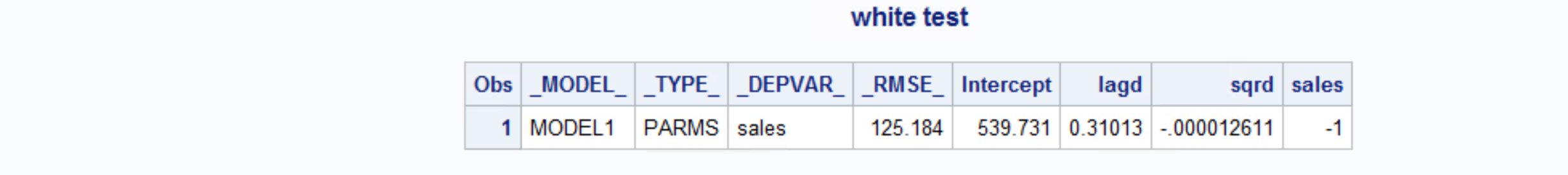
c = -0.0000126

The Bass Model is:

St = 539.73 + 0.310 Nt-1 -0.000126 (Nt-1)2

Regression output:





By solving for the bass model parameters from regression coefficients we get:

M= {-b-sqrt (b2 – 4ac)}/(2c)

p=a/M

q=p+b

substituting a, b, c values gives:

p=0.020581

q=0.33071

M=26225.01

Peak sales S\*

S\* = M\*(p+q)2 / (4\*q)

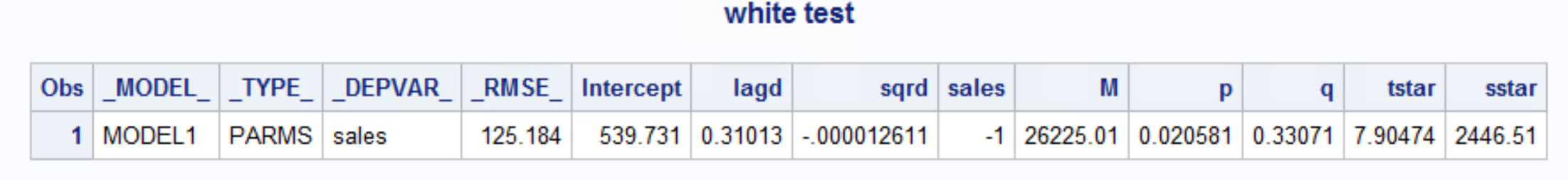
Peak sales = 2446.51

Time to reach peak sales t\*

t\* = ln (q/p)\*1/(p+q)

Time of occurrence = 7.9047

Bass Model Parameters output:



* 1. **Predict sales in each period using only the model parameters p, q, and M and the fact that sales at time period 0=0.**

The equation for forecast of sales using p, q, M is:

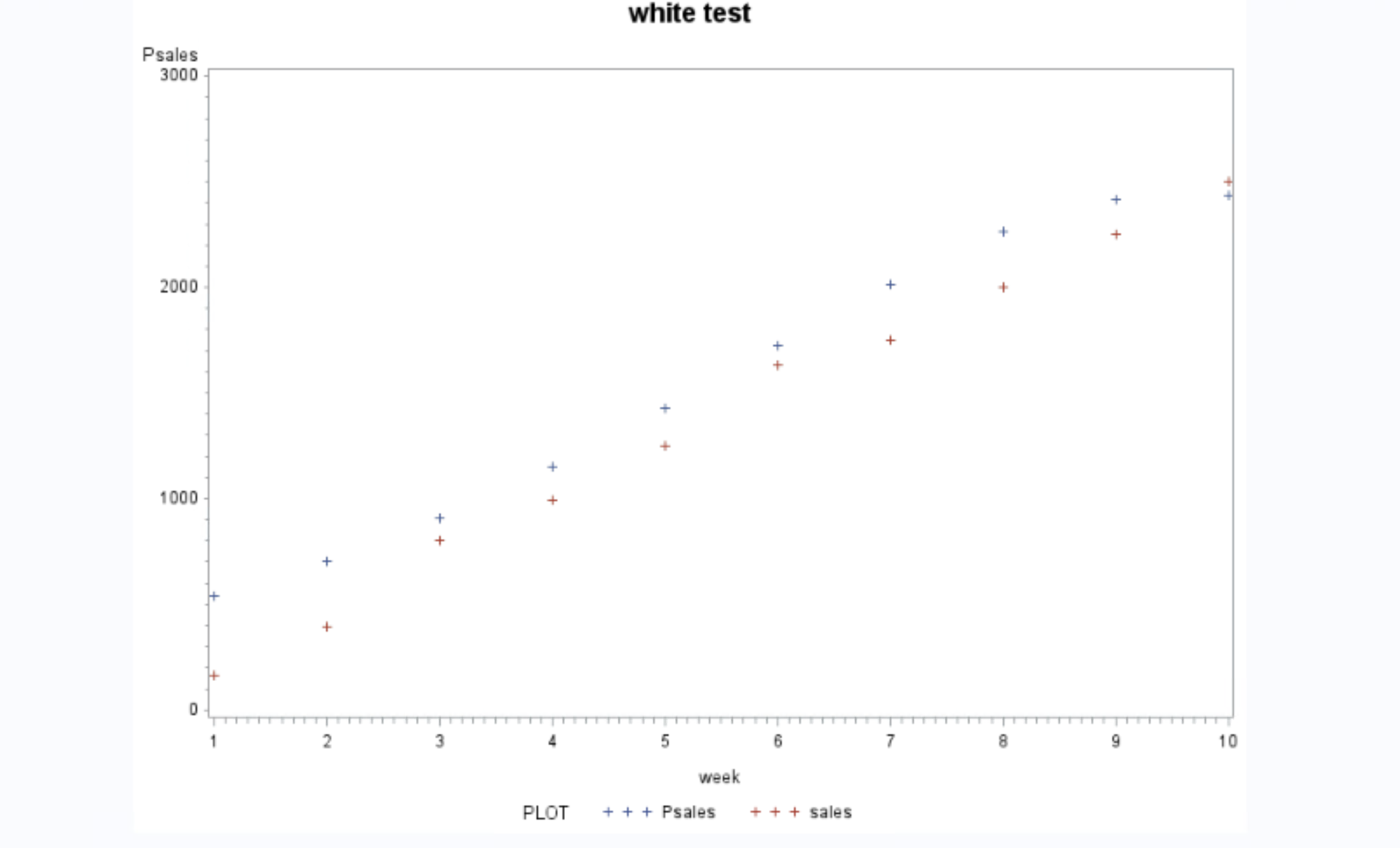
St = [p+(q/M)\*Nt-1]\* [M – Nt-1]

We get the predicted value of sales as:



* 1. **Plot a graph of actual versus predicted sales.**

proc gplot;plot psales\*week sales\*week/overlay;run;



We observe that predicted value of sales for initial weeks (1 and 2) is higher than the actual sales and the difference is large. Predicted sales is close to actual sales for week 4 to 6. The difference in the predicted and actual values again increases considerably from week 7. It is observed that the predicted value of sales for week 10 is less than the actual sales.

**Q.3 A conjoint study was undertaken by a detergent manufacturer. The attributes that were considered were**

**Brand (Complete, Smile, Wave)**

**Scent (fresh, lemon, Unscented)**

**Whether there was a softener or not (Y, N)**

**Size of packet (32, 48, 64)**

**Price (2.99, 3.99, 4.99)**

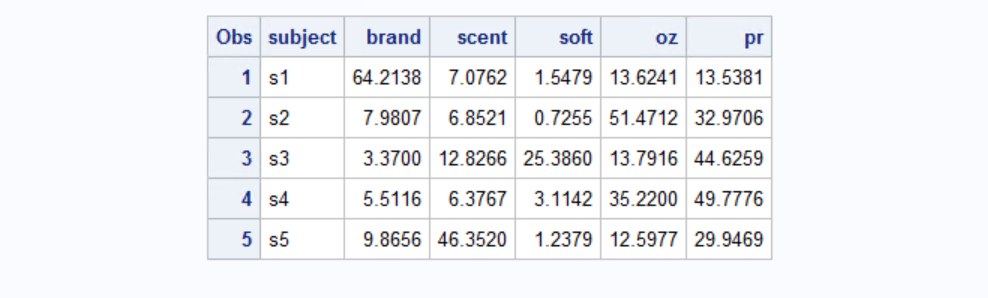
**The preferences of five respondents s1, s2, s3, s4, s5 were obtained for some combination of attributes on a 1-9 point scale with 9 indicating a higher preference.**

1. **Find the importance weights and part-worths for each respondent using PROC TRANSREG.**
2. **Predict the choice (using logit rule) for each respondent (s1-s5) for each of the following combinations using your estimates in question 1 above.**

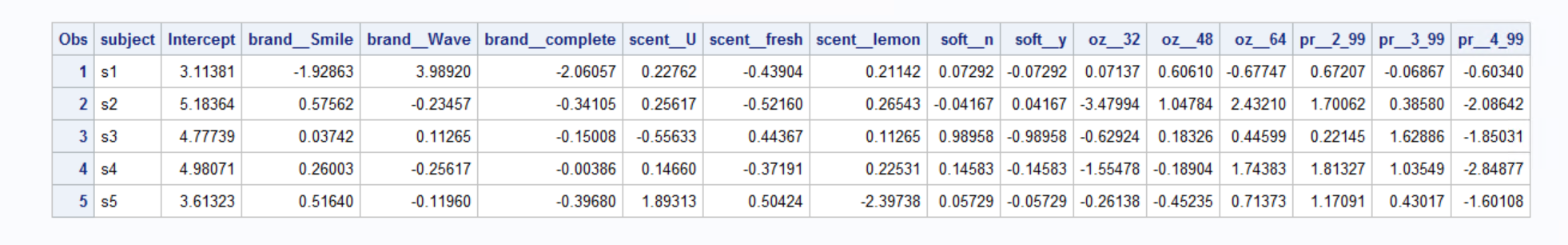
Conjoint analysis is used to study consumers’ product preferences and simulate consumer choice. Subjects provide data about their preferences for hypothetical products defined by attribute combinations. Conjoint analysis decomposes the judgment data into components, based on qualitative attributes of the products. A numerical part-worth utility value is computed for each level of each attribute.

PROC TRANSREG is used to perform a metric conjoint analysis. The output from the metric conjoint analysis is requested by specifying the utilities option in the proc statement. The value specified in the separators= option, i.e. a comma followed by a blank, is used in constructing the labels for the part-worth utilities in the displayed output. With these options, the labels consist of the class variable name, a comma, a blank and the values of the class variables.

The importance value is computed from the part-worth utility range for each factor (attribute). Each range is divided by the sum of all ranges and multiplied by 100. The factors with the largest part-worth utility ranges are the most important in determining preference. The below table shows important utility for each respondent.

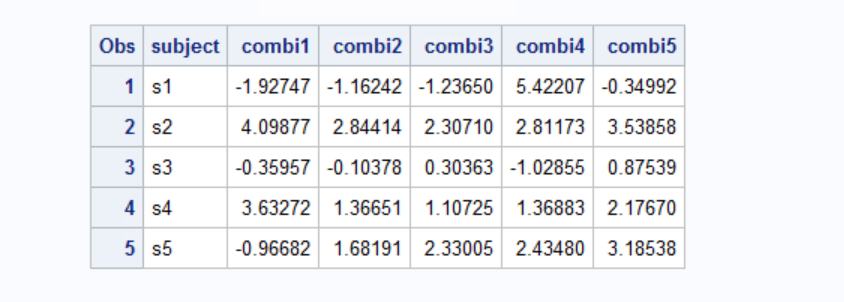


The part-worth utilities for the attribute levels are the parameter estimates. The below table shows part-worth for each attribute for each respondent.

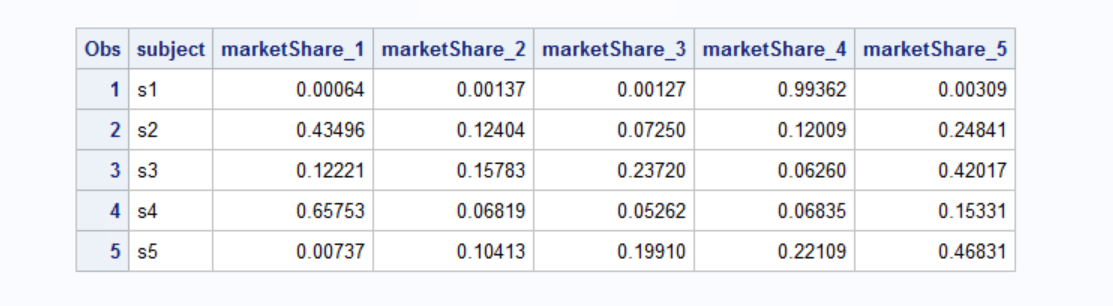


/\*3.2\*/

Utilities for the combinations mentioned for each respondent.



Market Share for the combinations for each respondent is calculated using Logit Rule The below table shows the market share of each combination for each respondent.



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | S1 | S2 | S3 | S4 | S5 |
| PR(1) | Complete | Lemon | Y | 64 | 2.99 | 0.00064 | 0.43496 | 0.12221 | 0.65753 | 0.00737 |
| Pr(2) | Smile | Fresh | Y | 48 | 2.99 | 0.00137 | 0.12404 | 0.15783 | 0.06819 | 0.10413 |
| Pr(3) | Smile | U | Y | 48 | 3.99 | 0.00127 | 0.07250 | 0.23720 | 0.05262 | 0.19910 |
| Pr(4) | Wave | U | Y | 48 | 2.99 | 0.99362 | 0.12009 | 0.05262 | 0.06835 | 0.22109 |
| Pr(5) | Smile | U | N | 48 | 2.99 | 0.00309 | 0.24841 | 0.42017 | 0.15331 | 0.46831 |
| **Choice** |  |  |  |  |  | **4** | **1** | **5** | **1** | **5** |

Q.3.2)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| complete | lemon | y | 64 | 2.99 |  |  |  |  |  |
| Smile | fresh | y | 48 | 2.99 |  |  |  |  |  |
| Smile | u | y | 48 | 3.99 |  |  |  |  |  |
| Wave | u | y | 48 | 2.99 |  |  |  |  |  |
| Smile | u | n | 48 | 2.99 |  |  |  |  |  |

Referring to the output from 3.1:

For S1:

1. complete, lemon, y, 64, 2.99

y = 3.1138 -2.0606 + 0.2114 - 0.0729 - 0.6775 + 0.6721 = 1.1863

1. Smile, fresh, y, 48, 2.99

y = 3.1138 -1.9286 - 0.4390 - 0.0729 + 0.6061 + 0.6721 = 1.9515

1. Smile, U, y, 48, 3.99

Y = 3.1138 -1.9286 + 0.2276 - 0.0729 + 0.6061 - 0.0687 = 1.8773

1. Wave, U, y, 48, 2.99

Y =3.1138 + 3.9892 + 0.2276 - 0.0729 + 0.6061 - 0.6721 = 7.1917

1. Smile, U, n, 48, 2.99

Y = 3.1138 -1.9286 + 0.2276 + 0.0729 + 0.6061 - 0.6721 = 1.4197

For S2:

1. complete, lemon, y, 64, 2.99

y = 5.1836 -0.3410 + 0.2654 + 0.0417 + 2.4321+ 1.7006 = 9.2824

1. Smile, fresh, y, 48, 2.99

y = 5.1836 + 0.5756 -0.5216 + 0.0417+ 1.0478 + 1.7006=8.0277

1. Smile, U, y, 48, 3.99

Y = 5.1836 + 0.5756 + 0.2562+ 0.0417+ 1.0478 + 0.3858= 7.49

1. Wave, U, y, 48, 2.99

Y = 5.1836 -0.2346 + 0.2562+ 0.0417+ 1.0478 + 1.7006 = 7.98

1. Smile, U, n, 48, 2.99

Y = 5.1836 + + 0.5756 + 0.2562 -0.0417 + 1.0478 + 1.7006 = 8.7

For S3:

1. complete, lemon, y, 64, 2.99

y = 4.7774 + 0.1127-0.150 -0.9896+ 0.4460+ 0.2215 = 4.411

1. Smile, fresh, y, 48, 2.99

y = 4.7774 + 0.0374 +0.4437 -0.9896 + 0.1833 + 0.2215 =4.66

1. Smile, U, y, 48, 3.99

Y = 4.7774 + 0.0374 -0.5563 -0.9896 + 0.1833 + 1.6289= 5.06

1. Wave, U, y, 48, 2.99

Y = 4.7774 +0.1127 -0.5563 + -0.9896 + 0.1833+ 0.2215 =3.75

1. Smile, U, n, 48, 2.99

Y = 4.7774 + 0.0374 -0.5563 + 0.9896 + 0.1833+ 0.2215 = 5.63

For S4:

1. complete, lemon, y, 64, 2.99

y = 4.9807 -0.0039 + 0.2253 -0.1458 + 1.7438 + 1.8133= 8.60

1. Smile, fresh, y, 48, 2.99

y = 4.9807+ 0.2600 -0.3719 -0.1458 -0.1890 + 1.8133=6.36

1. Smile, U, y, 48, 3.99

Y = 4.9807+ 0.2600 + 0.1466 -0.1458 -0.1890 +1.0355= 6.09

1. Wave, U, y, 48, 2.99

Y = 4.9807 -0.2562+ 0.1466 -0.1458 -0.1890+ 1.8133 =6.35

1. Smile, U, n, 48, 2.99

Y = = 4.9807+ 0.2600+ + 0.1466+0.1458-0.1890+ 1.8133 = 6.15

For S5:

1. complete, lemon, y, 64, 2.99

y = 3.6132 -0.3968 -2.3974 -0.0573 + 0.7137 + 1.1709 = 1.66

1. Smile, fresh, y, 48, 2.99

y = 3.6132 + 0.5164 + 0.5042 -0.0573 -0.4524 + 1.1709 = 5.29

1. Smile, U, y, 48, 3.99

Y = 3.6132+ 0.5164 +1.8931 -0.0573 -0.4524 +0.4302 =5.94

1. Wave, U, y, 48, 2.99

Y = 3.6132 -0.1196 +1.8931-0.0573-0.4524+ 1.1709 = 6.06

1. Smile, U, n, 48, 2.99

Y = 3.6132-0.1196 +1.8931+0.0573-0.4524+ 1.1709 =6.16